



Munich Personal RePEc Archive

Improving the Efficiency of GMM Estimators for Dynamic Panel Models

Ahmed H. Youssef and Ahmed A. El-Sheikh and Mohamed R. Abonazel

11. June 2014

Online at <https://mpa.ub.uni-muenchen.de/68675/>

MPRA Paper No. 68675, posted 8. January 2016 14:23 UTC

Improving the Efficiency of GMM Estimators for Dynamic Panel Models

Ahmed H. Youssef, Ahmed A. El-Sheikh, and Mohamed R. Abonazel

Department of Applied Statistics and Econometrics,
Institute of Statistical Studies and Research,
Cairo University, Egypt

Abstract

In dynamic panel models, the generalized method of moments (GMM) has been used in many applications since it gives efficient estimators. This efficiency is affected by the choice of the initial weight matrix. It is common practice to use the inverse of the moment matrix of the instruments as an initial weight matrix. However, an initial optimal weight matrix is not known, especially in the system GMM estimation procedure. Therefore, we present the optimal weight matrix for level GMM estimator, and suboptimal weight matrices for system GMM estimator, and use these matrices to increase the efficiency of GMM estimator. Using the Kantorovich inequality (KI), we find that the potential efficiency gain becomes large when the variance of individual effects increases compared to the variance of the errors.

Keywords: Dynamic panel data; Generalized method of moments; KI upper bound; Optimal and suboptimal weighting matrices

1. Introduction

An asymptotically efficient estimator can be obtained through the two-step procedure in the standard GMM estimation. In the first step, an initial positive semidefinite weight matrix is used to obtain consistent estimates of the parameters. Given these consistent estimates, a weight matrix can be constructed that is consistent for the efficient weight matrix, and this weight matrix is used for the asymptotically efficient two-step estimates. It is well known (see, *e.g.* Arellano and Bond, 1991) that the two-step estimated standard errors have a small-sample downward bias in this dynamic panel data setting, and one-step estimates with robust standard errors are often preferred. Although an efficient weight matrix for the difference model with errors that are homoskedastic and that are not serially correlated is easily derived, this is not the case for the system GMM estimator.

It is generally known that using many instruments can improve the efficiency of various GMM estimators (Arellano and Bover, 1995; Ahn and Schmidt, 1995; Blundell and Bond, 1998). Therefore, the system GMM estimator, which presented by Blundell and Bond (1998), is more efficient than first-difference GMM and level GMM estimators, which presented by

Arellano and Bond (1991) and Arellano and Bover (1995), respectively. Despite the substantial efficiency gain, using many instruments has two important drawbacks: increased bias and unreliable inference (Newey and Smith, 2004; Hayakawa, 2007). In this paper, we investigate how to decrease bias while increasing efficiency in the level and system GMM estimations by using the optimal (or suboptimal) weight matrices to obtain more efficient estimates of the parameters.

The remainder of this paper is organized as follows. Section 2 provides the model and reviews the conventional first-difference, level, and system GMM estimators. Section 3 presents the various optimal and suboptimal weight matrices for GMM estimators; we use KI upper bound to investigate the magnitude of the efficiency gain from use our suggested weight matrices. Finally, Section 4 offers the concluding remarks.

2. The Model and GMM Estimators

Consider a simple dynamic panel data process of the form

$$y_{it} = \phi y_{i,t-1} + \mu_i + \varepsilon_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (1)$$

Under the following assumptions:

- (i) ε_{it} are i.i.d across time and individuals and independent of μ_i and y_{i1} with $E(\varepsilon_{it}) = 0, Var(\varepsilon_{it}) = \sigma_\varepsilon^2$.
- (ii) μ_i are i.i.d across individuals with $E(\mu_i) = 0, Var(\mu_i) = \sigma_\mu^2$.
- (iii) The initial observations satisfy $y_{i1} = \frac{\mu_i}{1-\phi} + w_{i1}$ for $i = 1, \dots, N$, where $w_{i1} = \sum_{j=0}^{\infty} \phi^j \varepsilon_{i,1-j}$ and independent of μ_i .

Assumptions (i) and (ii) are the same as in Blundell and Bond (1998), while assumption (iii) has been developed by Alvarez and Arellano (2003).

Stacking equation (1) over time, we obtain

$$y_i = \phi y_{i,-1} + u_i, \quad (2)$$

where $y_i = (y_{i3}, \dots, y_{iT})', y_{i,-1} = (y_{i2}, \dots, y_{i,T-1})', u_i = (u_{i3}, \dots, u_{iT})'$, with $u_{it} = \mu_i + \varepsilon_{it}$.

Given these assumptions, we get three types of GMM estimators. These include first-difference GMM estimator, level GMM estimator and system GMM estimator.

2.1. First-difference GMM estimator

In model (2), the individual effect (μ_i) causes a severe correlation between the lagged endogenous variable ($y_{i,-1}$) and the error term (u_i). In order to eliminate the individual effect, Arellano and Bond (1991) take first differences of model (2):

$$\Delta y_i = \phi \Delta y_{i,-1} + \Delta u_i, \quad (3)$$

where $\Delta y_i = (y_{i3} - y_{i2}, \dots, y_{iT} - y_{i,T-1})'$, $\Delta y_{i,-1} = (y_{i2} - y_{i1}, \dots, y_{i,T-1} - y_{i,T-2})'$, and $\Delta u_i = (u_{i3} - u_{i2}, \dots, u_{iT} - u_{i,T-1})'$, and then showed that

$$E(H_i^{D'} \Delta u_i) = 0, \quad (4)$$

where

$$H_i^D = \begin{pmatrix} y_{i1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-2} \end{pmatrix}. \quad (5)$$

Using (4) as the orthogonal conditions in the GMM, Arellano and Bond (1991) constructed the one-step first-difference GMM (DIF) estimator for ϕ , which is given by

$$\hat{\phi}^D = (\Delta y'_{-1} H^D W_N^D H^{D'} \Delta y_{-1})^{-1} \Delta y'_{-1} H^D W_N^D H^{D'} \Delta y, \quad (6)$$

where $\Delta y_{-1} = (\Delta y'_{1,-1}, \dots, \Delta y'_{N,-1})'$, $\Delta y = (\Delta y'_1, \dots, \Delta y'_N)'$, $H^D = (H_1^{D'}, \dots, H_N^{D'})'$, and

$$W_N^D = \left(\frac{1}{N} \sum_{i=1}^N H_i^{D'} D H_i^D \right)^{-1}, \quad (7)$$

where $D = FF'$, and F is a $(T-2) \times (T-1)$ first-difference operator matrix

$$F = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}. \quad (8)$$

Blundell and Bond (1998) showed that when ϕ is close to unity and/or $\sigma_\mu^2/\sigma_\varepsilon^2$ increases the instruments matrix (5) becomes invalid. This means that the DIF estimator has the weak instruments problem.

2.2. Level GMM Estimator

Arellano and Bover (1995) suggested to eliminate the individual effect from instrumental variable, while, as mentioned above, Arellano and Bond (1991) proposed to eliminate it from the model. Explicitly, Arellano and Bover (1995) considered the level model (2) and then showed that the instrumental variable matrix

$$H_i^L = \begin{pmatrix} \Delta y_{i2} & 0 & \dots & 0 \\ 0 & \Delta y_{i3} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \Delta y_{i,T-1} \end{pmatrix}, \quad (9)$$

which not contains individual effect and satisfies the orthogonal conditions

$$E(H_i^{L'} u_i) = 0. \quad (10)$$

Using (10), Arellano and Bover's (1995) one-step level GMM (LEV) estimator is calculated as follows:

$$\hat{\phi}^L = (y'_{-1} H^L W_N^L H^{L'} y_{-1})^{-1} y'_{-1} H^L W_N^L H^{L'} y, \quad (11)$$

where $y_{-1} = (y'_{1,-1}, \dots, y'_{1,-N})'$, $y = (y'_1, \dots, y'_N)'$, $H^L = (H_1^{L'}, \dots, H_N^{L'})'$, and

$$W_N^L = \left(\frac{1}{N} \sum_{i=1}^N H_i^{L'} H_i^L \right)^{-1}. \quad (12)$$

2.3. System GMM Estimator

Blundell and Bond (1998) proposed a system GMM estimator in which the moment conditions of DIF and LEV are used jointly to avoid weak instruments and improve the efficiency of the estimator. The moment conditions used in constructing the system GMM estimator are given by

$$E(H_i^{S'} u_i^S) = 0, \quad (13)$$

where $u_i^S = (\Delta u'_i, u'_i)'$ and H_i^S is a $2(T-2) \times m$ block diagonal matrix given by

$$H_i^S = \begin{pmatrix} H_i^D & 0 \\ 0 & H_i^L \end{pmatrix}. \quad (14)$$

Using (13), the one-step system GMM (SYS) estimator is calculated as follows:

$$\hat{\phi}^S = (y_{-1}^{S'} H^S W_{N,G}^S H^{S'} y_{-1}^S)^{-1} y_{-1}^{S'} H^S W_{N,G}^S H^{S'} y^S, \quad (15)$$

where $y_{-1}^S = [(\Delta y'_{1,-1}, y'_{1,-1}), \dots, (\Delta y'_{N,-1}, y'_{N,-1})]'$, $y^S = [(\Delta y'_1, y'_1), \dots, (\Delta y'_N, y'_N)]'$, $H^S = (H_1^{S'}, \dots, H_N^{S'})'$, and

$$W_{N,G}^S = \left(\frac{1}{N} \sum_{i=1}^N H_i^{S'} G H_i^S \right)^{-1}, \quad (16)$$

where

$$G = \begin{pmatrix} D & 0 \\ 0 & I_{T-2} \end{pmatrix}. \quad (17)$$

Blundell and Bond (1998) used also the identity matrix (I_{2T-4}) instead to G :

$$W_{N,I}^S = \left(\frac{1}{N} \sum_{i=1}^N H_i^{S'} H_i^S \right)^{-1}, \quad (18)$$

in their first step of two-step system GMM estimator, which yields the simple system GMM estimator. This is certainly not optimal either, but is easy and could perhaps suit well as first step in a two-step procedure.

3. The Optimal and Suboptimal Weighting Matrices

Generally, using the moment conditions, the GMM estimator $\hat{\phi}$ for ϕ_0 minimizes

$$\left[\frac{1}{N} \sum_{i=1}^N f_i(\phi) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N f_i(\phi) \right], \quad (19)$$

with respect to ϕ , where W_N is a positive semidefinite weight matrix which satisfies $p\lim_{N \rightarrow \infty} W_N = W$, with W a positive definite matrix. Regularity conditions are in place such that $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N f_i(\phi) = E\{f(\phi)\}$ and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N f_i(\phi_0) \rightarrow N(0, \Psi). \quad (20)$$

Let $F(\phi) = E(\partial f_i(\phi)/\partial \phi)$ and $F_0 \equiv F(\phi_0)$, then $\sqrt{N}(\hat{\phi} - \phi_0)$ has a limiting normal distribution, $\sqrt{N}(\hat{\phi} - \phi_0) \rightarrow N(0, V_W)$, where

$$V_W = (F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1}. \quad (21)$$

As is clear from the expression of the asymptotic variance matrix V_W , in (21), the efficiency of the GMM estimator is affected by the choice of the weight matrix W_N . An optimal choice is a weight matrix for which $W = \Psi^{-1}$. The asymptotic variance matrix is then given by $(F_0' \Psi^{-1} F_0)^{-1}$. For any other W , the GMM estimator is less efficient as

$$(F_0' \Psi^{-1} F_0)^{-1} \leq (F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1}. \quad (22)$$

To assess the potential loss in efficiency from using this initial weight matrix, the following expression for the upper bound of the efficiency loss has been derived by Liu and Neudecker (1997) on the basis of the Kantorovich inequality:

$$(F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1} \leq \frac{(\lambda_1 + \lambda_p)^2}{4\lambda_1 \lambda_p} (F_0' \Psi^{-1} F_0)^{-1}, \quad (23)$$

and the upper bound is $B_{KI} = \frac{(\lambda_1 + \lambda_p)^2}{4\lambda_1\lambda_p}$, where $\lambda_i > 0$; $i = 1, \dots, p$ are the eigenvalues of the $p \times p$ matrix ΨW . If B_{KI} increased led to a further loss of efficiency.

In first-difference GMM estimation, Kiviet (2007) showed that W_N^D matrix is the optimal weight matrix for DIF estimator when $\varepsilon_i \sim i.i.d. (0, \sigma_\varepsilon^2 I_T)$. So, we will not adjust this estimator by new weight matrix because it will not improve the efficiency of the estimation.

Lemma 1: Let assumptions (i) to (iii) hold, and under the orthogonal conditions in (10). The optimal weight matrix for LEV estimator is given by

$$W_N^{OL} = \left(\frac{1}{N} \sum_{i=1}^N H_i^{L'} J_{T-2} H_i^L \right)^{-1}, \quad (24)$$

where

$$J_{T-2} = \begin{pmatrix} 1 + \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 + \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 + \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 + \rho \end{pmatrix}; \rho = \sigma_\mu^2 / \sigma_\varepsilon^2. \quad (25)$$

Proof: For $t = 2, 3, \dots, T-1$, there are $T-2$ over-identifying moment conditions, and Ψ matrix is given by

$$\begin{aligned} \Psi &= p\lim \frac{1}{N} \sum_{i=1}^N (H_i^{L'} \hat{u}_i \hat{u}_i' H_i^L) \\ &= p\lim \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} (\Delta y_{i2})^2 (\mu_i + \varepsilon_{i3})^2 & \Delta y_{i2} \Delta y_{i3} (\mu_i + \varepsilon_{i3})(\mu_i + \varepsilon_{i4}) & \cdots & \Delta y_{i2} \Delta y_{i,T-1} (\mu_i + \varepsilon_{i3})(\mu_i + \varepsilon_{iT}) \\ \Delta y_{i2} \Delta y_{i3} (\mu_i + \varepsilon_{i3})(\mu_i + \varepsilon_{i4}) & (\Delta y_{i3})^2 (\mu_i + \varepsilon_{i4})^2 & \cdots & \Delta y_{i3} \Delta y_{i,T-1} (\mu_i + \varepsilon_{i4})(\mu_i + \varepsilon_{iT}) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta y_{i2} \Delta y_{i,T-1} (\mu_i + \varepsilon_{i3})(\mu_i + \varepsilon_{iT}) & \Delta y_{i3} \Delta y_{i,T-1} (\mu_i + \varepsilon_{i4})(\mu_i + \varepsilon_{iT}) & \cdots & (\Delta y_{i,T-1})^2 (\mu_i + \varepsilon_{iT})^2 \end{bmatrix} \\ \Psi &= \begin{bmatrix} (\sigma_\mu^2 + \sigma_\varepsilon^2) E(\Delta y_{i2})^2 & \sigma_\mu^2 E(\Delta y_{i2} \Delta y_{i3}) & \cdots & \sigma_\mu^2 E(\Delta y_{i2} \Delta y_{i,T-1}) \\ \sigma_\mu^2 E(\Delta y_{i2} \Delta y_{i3}) & (\sigma_\mu^2 + \sigma_\varepsilon^2) E(\Delta y_{i3})^2 & \cdots & \sigma_\mu^2 E(\Delta y_{i3} \Delta y_{i,T-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 E(\Delta y_{i2} \Delta y_{i,T-1}) & \sigma_\mu^2 E(\Delta y_{i3} \Delta y_{i,T-1}) & \cdots & (\sigma_\mu^2 + \sigma_\varepsilon^2) E(\Delta y_{i,T-1})^2 \end{bmatrix}. \end{aligned}$$

We use an initial weight matrix equal to J_{T-2} , so W in this case is obtained by

$$\begin{aligned} W &= \left(p\lim \frac{1}{N} \sum_{i=1}^N H_i^{L'} J_{T-2} H_i^L \right)^{-1} \\ &= \left(p\lim \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} (1 + \rho)(\Delta y_{i2})^2 & \rho \Delta y_{i2} \Delta y_{i3} & \cdots & \rho \Delta y_{i2} \Delta y_{i,T-1} \\ \rho \Delta y_{i2} \Delta y_{i3} & (1 + \rho)(\Delta y_{i3})^2 & \cdots & \rho \Delta y_{i3} \Delta y_{i,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho \Delta y_{i2} \Delta y_{i,T-1} & \rho \Delta y_{i3} \Delta y_{i,T-1} & \cdots & (1 + \rho)(\Delta y_{i,T-1})^2 \end{bmatrix} \right)^{-1} \end{aligned}$$

$$W = \begin{bmatrix} (1 + \rho)E(\Delta y_{i2})^2 & \rho E(\Delta y_{i2}\Delta y_{i3}) & \cdots & \rho E(\Delta y_{i2}\Delta y_{iT-1}) \\ \rho E(\Delta y_{i2}\Delta y_{i3}) & (1 + \rho)E(\Delta y_{i3})^2 & \cdots & \rho E(\Delta y_{i3}\Delta y_{iT-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho E(\Delta y_{i2}\Delta y_{iT-1}) & \rho E(\Delta y_{i3}\Delta y_{iT-1}) & \cdots & (1 + \rho)E(\Delta y_{iT-1})^2 \end{bmatrix}^{-1}$$

We can rewrite W as:

$$W = \left(\frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (\sigma_\mu^2 + \sigma_\varepsilon^2)E(\Delta y_{i2})^2 & \sigma_\mu^2 E(\Delta y_{i2}\Delta y_{i3}) & \cdots & \sigma_\mu^2 E(\Delta y_{i2}\Delta y_{iT-1}) \\ \sigma_\mu^2 E(\Delta y_{i2}\Delta y_{i3}) & (\sigma_\mu^2 + \sigma_\varepsilon^2)E(\Delta y_{i3})^2 & \cdots & \sigma_\mu^2 E(\Delta y_{i3}\Delta y_{iT-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 E(\Delta y_{i2}\Delta y_{iT-1}) & \sigma_\mu^2 E(\Delta y_{i3}\Delta y_{iT-1}) & \cdots & (\sigma_\mu^2 + \sigma_\varepsilon^2)E(\Delta y_{iT-1})^2 \end{bmatrix} \right)^{-1}$$

$$W = \sigma_\varepsilon^2 \Psi^{-1}$$

$$\Psi W = \sigma_\varepsilon^2 (\Psi \Psi^{-1}) = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I_{T-2}.$$

Since the eigenvalues of ΨW matrix:

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{T-2} = \sigma_\varepsilon^2$$

and

$$B_{KI} = \frac{(\lambda_1 + \lambda_{T-2})^2}{4\lambda_1\lambda_{T-2}} = \frac{(2\sigma_\varepsilon^2)^2}{4\sigma_\varepsilon^4} = 1.$$

Therefore W_N^{OL} is the optimal weight matrix for level GMM estimator.

Note that the use of the optimal weight matrix W_N^{OL} can be described as inducing cross-sectional heterogeneity through ρ , and also can be explained as partially adopting a procedure of generalized least squares (GLS) to the level estimation, which is not done in Arellano and Bover (1995). So, using J_{T-2} , instead of I_{T-2} , certainly improve the efficiency of LEV estimator.

Remark 1: We find that the weight matrix W_N^L , given in (12), is the optimal weight matrix for level GMM estimator only in case of $\sigma_\mu^2 = 0$.

Figure 1 presents the efficiency bounds, B_{KI} , for LEV GMM estimator with the optimal, W_N^{OL} , and the conventional, W_N^L , weight matrices, when $T = 4$ and for various values of ϕ and ρ . We choose $\rho = 0, 0.5, 1, 5, 10$ and 25 to investigate the loss of efficiency when using W_N^L . Figure 1 showed that, when ϕ is small (close to zero) and/or ρ is large the efficiency loss could be quite severe when using W_N^L .

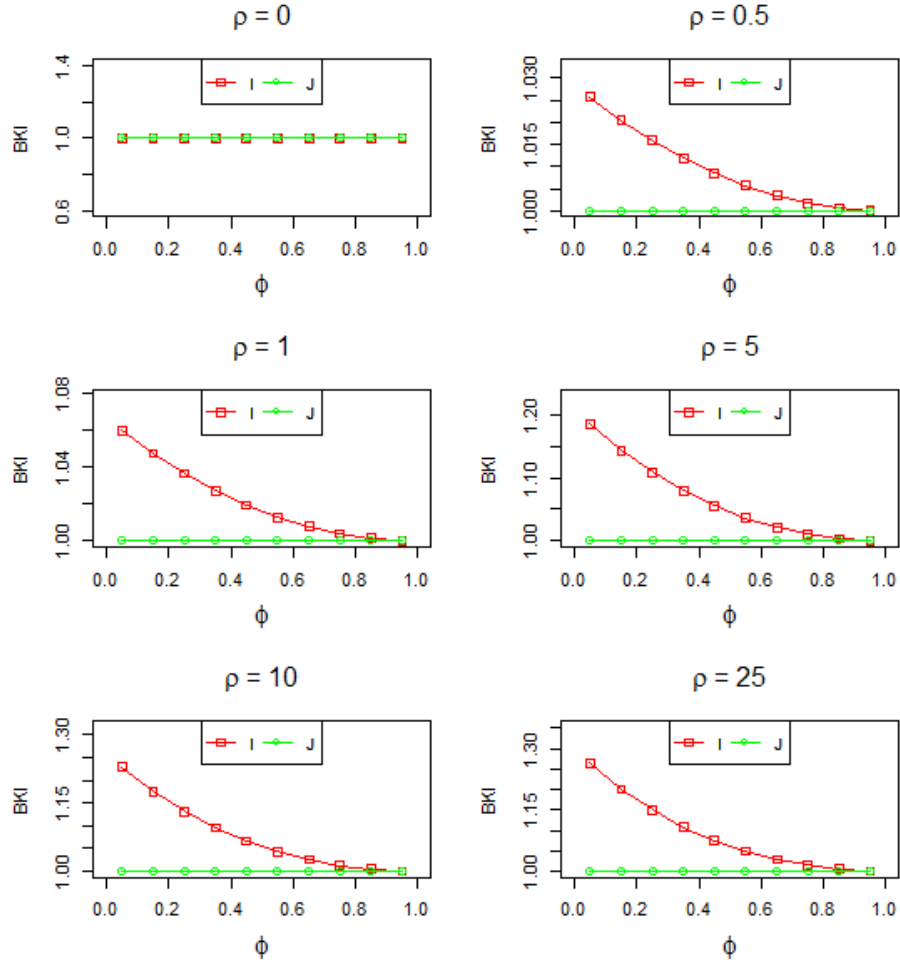


Figure 1: KI efficiency bounds for various weight matrices for LEV GMM estimator, when $T = 4$

In system GMM estimation, Windmeijer (2000) showed that the optimal weight matrix, for SYS estimator has only been obtained in case of $\sigma_\mu^2 = 0$, and this matrix is given by:

$$W_{N,c}^S = \left(\frac{1}{N} \sum_{i=1}^N H_i^{S'} G_c H_i^S \right)^{-1}, \quad (26)$$

where

$$G_c = \begin{pmatrix} D & C \\ C' & I_{T-2} \end{pmatrix}, \quad (27)$$

and C is a $(T - 2)$ square matrix:

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \ddots & \vdots \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \quad (28)$$

Lemma 2: Let assumptions (i) to (iii) hold, and under the orthogonal conditions in (13). If $\rho > 1$, we have that, the following suboptimal weight matrices:

$$W_{N,cj}^S = \left(\frac{1}{N} \sum_{i=1}^N H_i^{S'} G_{cj} H_i^S \right)^{-1}, \quad \text{with } G_{cj} = \begin{pmatrix} D & C \\ C' & J_{T-2} \end{pmatrix},$$

$$W_{N,j}^S = \left(\frac{1}{N} \sum_{i=1}^N H_i^{S'} G_j H_i^S \right)^{-1}, \quad \text{with } G_j = \begin{pmatrix} D & 0 \\ 0 & J_{T-2} \end{pmatrix},$$

are more efficiency, for SYS estimator, than the initial weight matrices $W_{N,G}^S$ and $W_{N,I}^S$.

Note. For simplicity, the lemma will be proved for $T = 3$.

Proof: Since $T = 3$, there are two over-identifying moment conditions:

$$E[y_{i1}(\Delta y_{i3} - \phi \Delta y_{i2})] = 0; E[\Delta y_{i2}(y_{i3} - \phi y_{i2})] = 0,$$

and Ψ matrix is given by

$$\Psi = \sigma_\varepsilon^2 \begin{bmatrix} 2\sigma_y^2 & -(1-\phi)\sigma_y^2 \\ -(1-\phi)\sigma_y^2 & 2\frac{\sigma_\mu^2 + \sigma_\varepsilon^2}{1+\phi} \end{bmatrix},$$

where

$$\sigma_y^2 = \frac{\sigma_\mu^2}{(1-\phi)^2} + \frac{\sigma_\varepsilon^2}{1-\phi^2}.$$

(a) In case of $W_{N,I}^S$, the W matrix is obtained by

$$W = \left(p\lim \frac{1}{N} \sum_{i=1}^N H_i^{S'} H_i^S \right)^{-1} = \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & 2\frac{\sigma_\varepsilon^2}{1+\phi} \end{bmatrix}^{-1},$$

and ΨW is given by

$$\Psi W = \begin{bmatrix} 2\sigma_\varepsilon^2 & -\frac{(1-\phi^2)\sigma_y^2}{2} \\ -(1-\phi)\sigma_\varepsilon^2 & \sigma_\mu^2 + \sigma_\varepsilon^2 \end{bmatrix}.$$

Since the eigenvalues of ΨW :

$$\lambda = \frac{1}{2} \left[(\sigma_\mu^2 + 3\sigma_\varepsilon^2) \pm \sqrt{\sigma_\varepsilon^2 + \sigma_\mu^2 - 2\sigma_\mu^2\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2\sigma_y^2(1-\phi)(1-\phi^2)} \right]$$

and

$$B_{KI,I} = \frac{(\sigma_\mu^2 + 3\sigma_\varepsilon^2)^2}{2\sigma_\varepsilon^2 [4(\sigma_\mu^2 + \sigma_\varepsilon^2) - \sigma_y^2(1-\phi)(1-\phi^2)]}. \quad (29)$$

(b) In case of $W_{N,G}^S$, the W matrix is obtained by

$$W = \left(p\lim \frac{1}{N} \sum_{i=1}^N H_i^{S'} G H_i^S \right)^{-1} = \begin{bmatrix} 2\sigma_y^2 & 0 \\ 0 & 2\frac{\sigma_\varepsilon^2}{1+\phi} \end{bmatrix}^{-1},$$

and ΨW is given by

$$\Psi W = \begin{bmatrix} \sigma_\varepsilon^2 & -\frac{(1-\phi^2)\sigma_y^2}{2} \\ \frac{-(1-\phi)\sigma_\varepsilon^2}{2} & \sigma_\mu^2 + \sigma_\varepsilon^2 \end{bmatrix}.$$

Then:

$$B_{KI,G} = \frac{(2\sigma_\varepsilon^2 + \sigma_\mu^2)^2}{\sigma_\varepsilon^2 [4(\sigma_\mu^2 + \sigma_\varepsilon^2) - \sigma_y^2(1-\phi)^2(1+\phi)]}. \quad (30)$$

(c) In case of $W_{N,j}^S$, the W matrix is obtained by

$$W = \left(p\lim \frac{1}{N} \sum_{i=1}^N H_i^{S'} G_j H_i^S \right)^{-1} = \begin{bmatrix} 2\sigma_y^2 & 0 \\ 0 & 2\frac{\sigma_\varepsilon^2(1+\rho)}{1+\phi} \end{bmatrix}^{-1},$$

and ΨW is given by

$$\Psi W = \begin{bmatrix} \sigma_\varepsilon^2 & -\frac{(1-\phi^2)\sigma_y^2}{2(1+\rho)} \\ \frac{-(1-\phi)\sigma_\varepsilon^2}{2} & \frac{\sigma_\mu^2 + \sigma_\varepsilon^2}{(1+\rho)} \end{bmatrix}.$$

Then:

$$B_{KI,G_j} = \frac{[\sigma_\varepsilon^2(2+\rho) + \sigma_\mu^2]^2}{(1+\rho)[4\sigma_\varepsilon^2(\sigma_\mu^2 + \sigma_\varepsilon^2) - \sigma_\varepsilon^2\sigma_y^2(1-\phi)^2(1+\phi)]}. \quad (31)$$

(d) In case of $W_{N,cj}^S$, the W matrix is obtained by

$$W = \left(p\lim \frac{1}{N} \sum_{i=1}^N H_i^{S'} G_{cj} H_i^S \right)^{-1} = \begin{bmatrix} 2\sigma_y^2 & -\frac{\sigma_\varepsilon^2}{1+\phi} \\ -\frac{\sigma_\varepsilon^2}{1+\phi} & 2\frac{\sigma_\varepsilon^2(1+\rho)}{1+\phi} \end{bmatrix}^{-1},$$

and ΨW is given by

$$\Psi W = \frac{2}{M} \begin{pmatrix} 2f - \frac{b}{2} & a - g \\ e - d & 2c - \frac{b}{2} \end{pmatrix}.$$

where

$$a = (1+\phi)\sigma_\varepsilon^2\sigma_y^2; b = (1-\phi^2)\sigma_\varepsilon^2\sigma_y^2; c = (1+\phi)(\sigma_\mu^2 + \sigma_\varepsilon^2)\sigma_y^2;$$

$$d = (1 + \rho).b; e = (\sigma_\mu^2 + \sigma_\varepsilon^2)\sigma_\varepsilon^2; f = (1 + \rho).a;$$

$$g = (1 + \phi)^2(1 - \phi)\sigma_y^4; M = 4(1 + \rho)(1 + \phi)\sigma_y^2 - \sigma_\varepsilon^2.$$

Then:

$$B_{KI,G_{cj}} = \frac{(2f + 2c - b)^2}{4 \left(4c.f - b.f - b.c + \frac{1}{4}b^2 - a.e + a.d + e.g - d.g \right)}. \quad (32)$$

From (29) to (32) we can conclude that, when $\rho > 1$, value of the efficiency bound B_{KI,G_j} less than value of $B_{KI,G}$ and $B_{KI,I}$. As well as, value of $B_{KI,G_{cj}}$ less than value of $B_{KI,G}$ and $B_{KI,I}$. Therefore, the suboptimal weight matrices $W_{N,j}^S$ and $W_{N,cj}^S$ are more efficiency than the weight matrices $W_{N,G}^S$ and $W_{N,I}^S$.

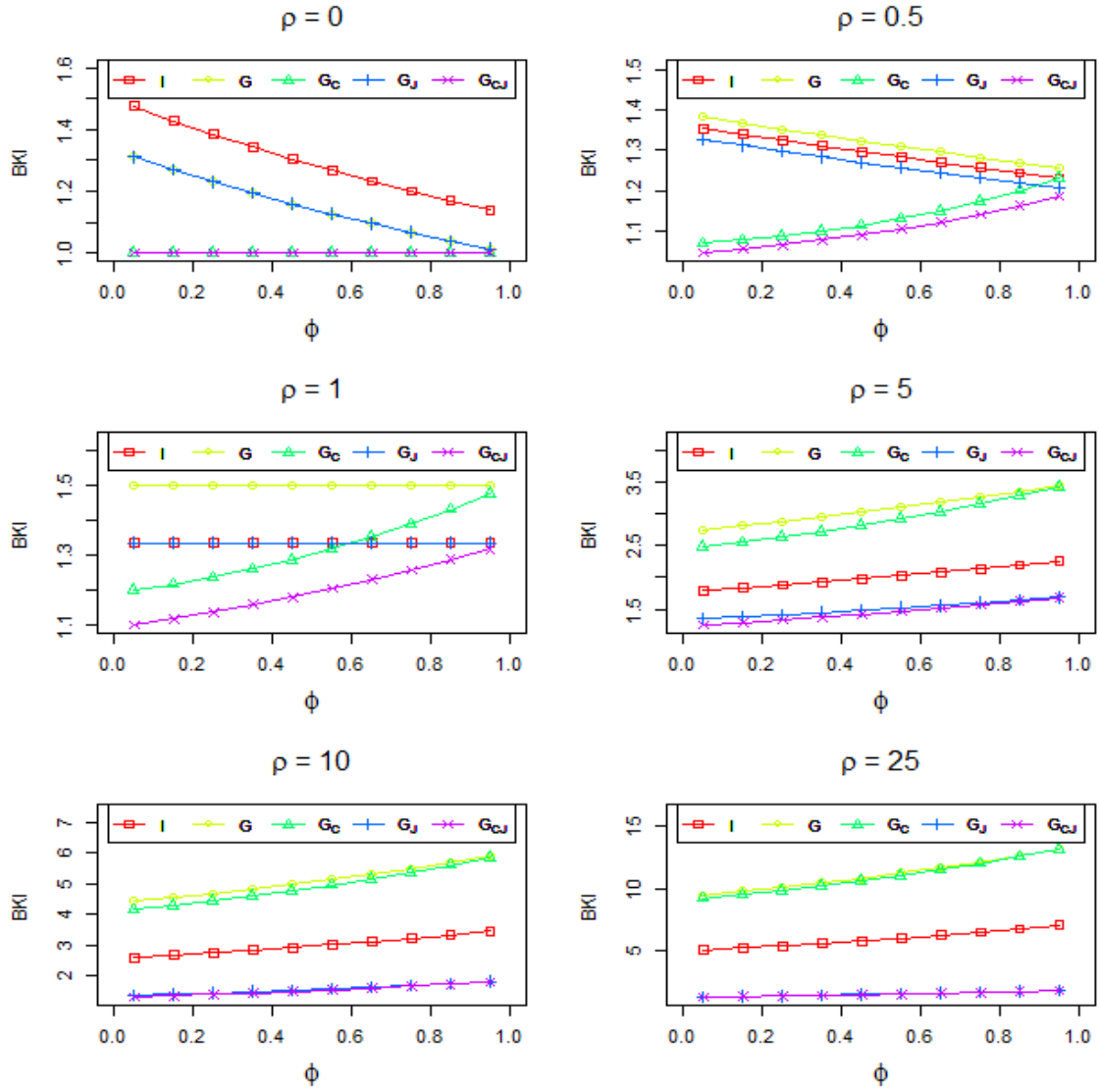


Figure 2: KI efficiency bounds for various weight matrices for SYS GMM estimator, when $T = 3$

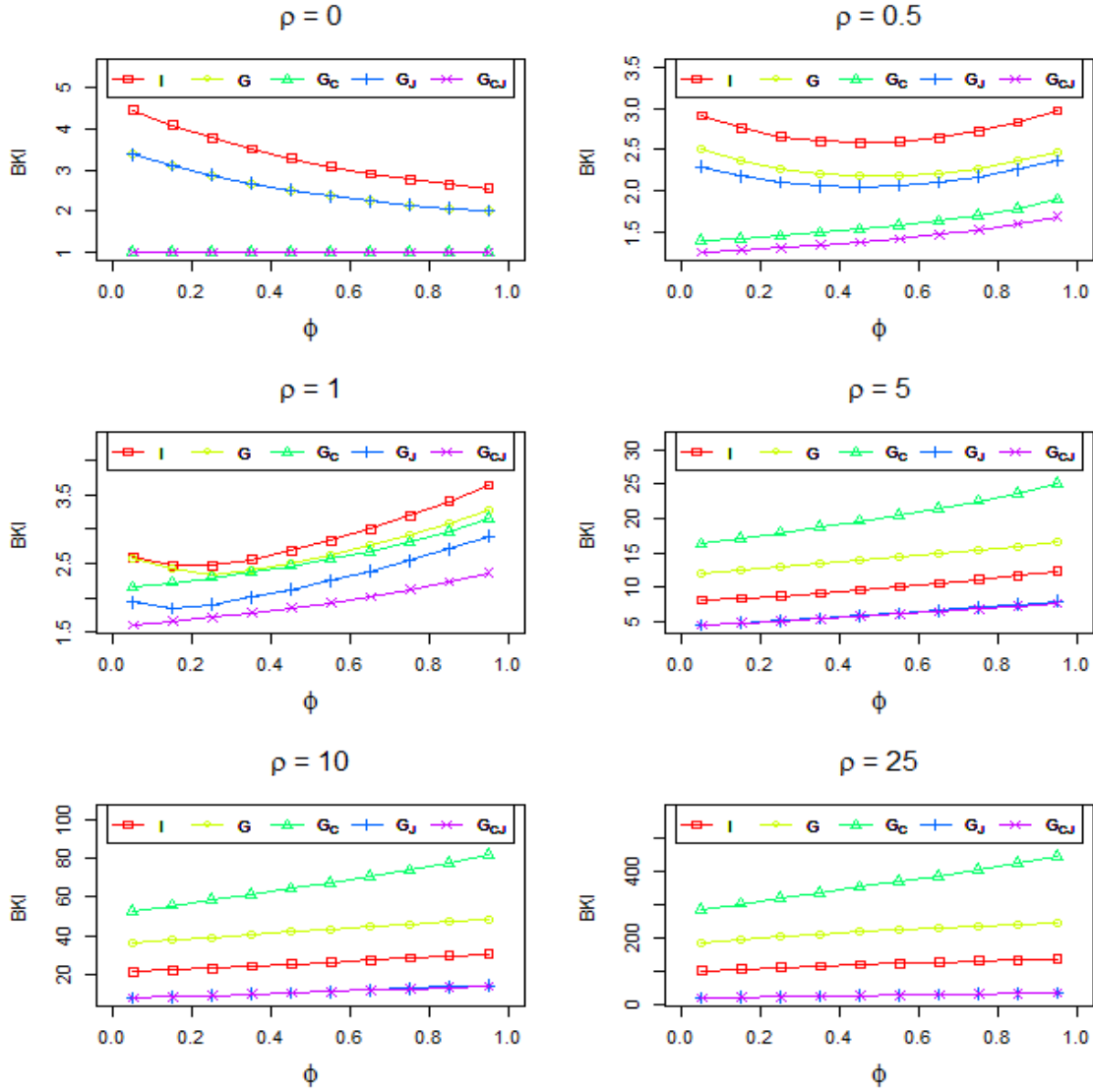


Figure 3: KI efficiency bounds for various weight matrices for SYS GMM estimator, when $T = 4$

Remark 2:

- (i) Although we consider the case of $T = 3$ in this paper, from the proof of Lemma 2 we find that what is stated in Lemma 2 still holds when $T > 3$ and this case is left for future research.
- (ii) When $T = 3$ and $\sigma_\mu^2 = \sigma_\varepsilon^2 = 1$, We find that $B_{KI,I} = B_{KI,G_j} = 4/3$, this means that the asymptotic variance of the SYS estimators, whether using $W_{N,I}^S$ or $W_{N,j}^S$, could potentially be 33% larger than the efficient estimator. Moreover, this percent will increase when $T > 3$ because of the value of B_{KI} increases with T as the number of moment restrictions increases.

Remark 3: When $\sigma_\mu^2 = 0$, we find that:

- (iii) $B_{KI,G_{cj}} = 1$, then the weight matrix $W_{N,cj}^S$ is the optimal weight matrix for SYS GMM estimator in this case, because it equivalent to $W_{N,c}^S$ matrix.
- (iv) $B_{KI,G} = B_{KI,G_j}$ for all T . Moreover, if $T = 3$, we get that $B_{KI,G} = B_{KI,G_j} = 4/(\phi + 3)$.

Figure 2 presents the efficiency bounds, B_{KI} , for SYS GMM estimator with the various weight matrices, when $T = 3$ and for various values of ϕ and ρ . We choose $\rho = 0, 0.5, 1, 5, 10$ and 25 to investigate the efficiency of various weight matrices. This Figure illustrates that, when $\rho > 1$ the values of B_{KI,G_j} and $B_{KI,G_{cj}}$ are the least, always, for all values of ϕ . Therefore, $W_{N,j}^S$ and $W_{N,cj}^S$ matrices are more efficiency. While Figure 3 presents the efficiency bounds when $T = 4$. The values of B_{KI} when $T = 4$ are larger than the values of B_{KI} when $T = 3$. Also, $W_{N,j}^S$ and $W_{N,cj}^S$ matrices are still the best efficiency.

4. Conclusion

The efficiency of GMM estimators is affected by the choice of the weight matrix. It is common practice to use the inverse of the moment matrix of the instruments as an initial weight matrix. This procedure led to a loss of efficiency. Therefore, we presented the optimal weight matrix, W_N^{OL} , for level GMM estimator. The KI efficiency bound showed that the efficiency loss could be quite severe when using W_N^L in level GMM estimator.

We presented suboptimal weight matrices, $W_{N,j}^S$ and $W_{N,cj}^S$, for system GMM estimator to improve the efficiency of GMM estimators. The KI efficiency bound showed that the $W_{N,j}^S$ and $W_{N,cj}^S$ matrices are more efficiency than the conventional weight matrices, $W_{N,G}^S$ and $W_{N,I}^S$, and the efficiency loss could be quite severe when using $W_{N,G}^S$ and $W_{N,I}^S$, especially when T gets large. When ρ is small, an efficiency gain can be made by using a weight matrix $W_{N,c}^S$ that is optimal under the assumption that $\sigma_\mu^2 = 0$. Consequently, we conclude that the $W_{N,j}^S$ and $W_{N,cj}^S$ matrices will provide more efficiency system GMM estimator for the practitioner. But in practice, ρ is unknown. So, we suggest use Jung and Kwon's (2007) estimator for ρ .

References

- [1] Ahn, S.C., and P. Schmidt, (1995). Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.
- [2] Alvarez, J., and M. Arellano, (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators, *Econometrica*, 71, 1121-1159.
- [3] Arellano, M., and S. Bond, (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-98.
- [4] Arellano, M., and O. Bover, (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

- [5] Blundell, R., and S. Bond, (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.
- [6] Hayakawa, K. (2007). Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models, *Economics Letters*, 95, 32-38.
- [7] Jung, H., and H. Kwon, (2007). An Alternative System GMM Estimation in Dynamic Panel Models, Hi-Stat Discussion Paper No. 217, Hitotsubashi University.
- [8] Kiviet, J.F. (2007). Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models. In *The Refinement of Econometric Estimation and Test Procedures*, ed. by G.D.A. Phillips and E. Tzavalis, Cambridge University Press, Cambridge, UK.
- [9] Liu, S., and H. Neudecker, (1997), Kantorovich Inequalities and Efficiency Comparisons for Several Classes of Estimators in Linear Models, *Statistica Neerlandica*, 51, 345-355.
- [10] Newey, W., and R. Smith, (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica*, 72, 219-255.
- [11] Windmeijer, F., (2000). Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models. In *Innovations in Multivariate Statistical Analysis*, ed. by R.D.H. Heijmans, D.S.G. Pollock and A. Satorra, A Festschrift for Heinz Neudecker. Advanced Studies in Theoretical and Applied Econometrics, vol. 36, Dordrecht: Kluwer Academic Publishers (IFS working paper W98/1).